Stat 102C - Lab 1

Levon Demirdjian

March 30 2016

In today's lab, we will be using Monte Carlo methods to approximate one-dimensional integrals. This will be a good review of what you have learned in your probability course. To start, consider the following nasty function:

$$g(x) = \frac{6\pi x^4 \exp(-x^2)(\log(x) + 3^x)}{(x+6)^{1/2}}. \tag{1}$$

Let's say our goal is to compute the integral, from 0 to 1, of this nasty function:

$$\int_0^1 \frac{6\pi x^4 \exp(-x^2)(\log(x) + 3^x)}{(x+6)^{1/2}} dx. \tag{2}$$

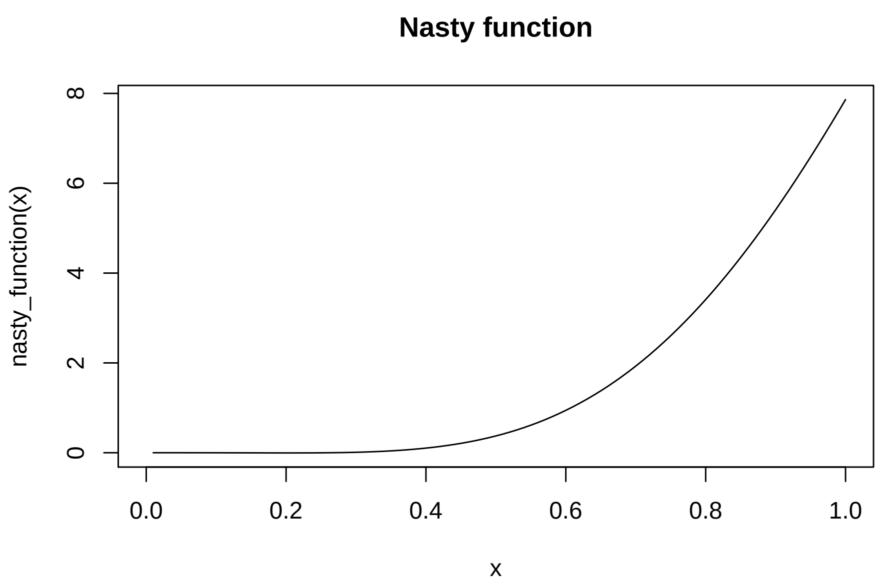
Of course, we can use R's built in integration function to do this:

```
nasty_function <- function(x){6*pi * (x^4)*exp(-x^2)*(log(x) + 3^x)/((x+6)^(0.5))}
integrate(nasty_function, lower = 0, upper = 1)
```

```
## 1.588643 with absolute error < 2.2e-10
```

Let's take a look at the plot of this nasty function:

```
curve(nasty_function, 0, 1, main = "Nasty function")
```



From calculus, we can solve integral (2) via . . . substitution? Or integration by parts? Or something else? Let me know if you find out how!

Problem 1

Consider an alternative scheme, where we use random sampling to approximate this integral with arbitrary precision. Let $X_1, ..., X_n \overset{iid}{\sim} U(0,1)$, where $U(0,1)$ is the uniform distribution on $(0,1)$. By the law of large numbers, we know that

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{p} E_U[g(X)],$$

where E_U is the expectation with respect to the uniform(0,1) distribution. Notice that

$$E_U[g(X)] = \int_0^1 g(x)1dx$$

is exactly the integral we are trying to approximate (I put the 1 inside of the integral to remind you that the pdf of a uniform(0,1) random variable is 1)! Therefore we have found a simple scheme to approximate this integral:

1. Step 1: Generate $X_1, ..., X_n \overset{iid}{\sim} U(0,1)$ with large n.
2. Step 2: Compute $g(X_i)$ for each X_i.
3. Step 3: Compute the mean of the $g(X_i)$'s. The mean will give us an approximation to the integral we seek.

The corresponding code is very simple:

```
n <- 10000
u <- runif(n, min = 0, max = 1)
mean(nasty_function(u))
```

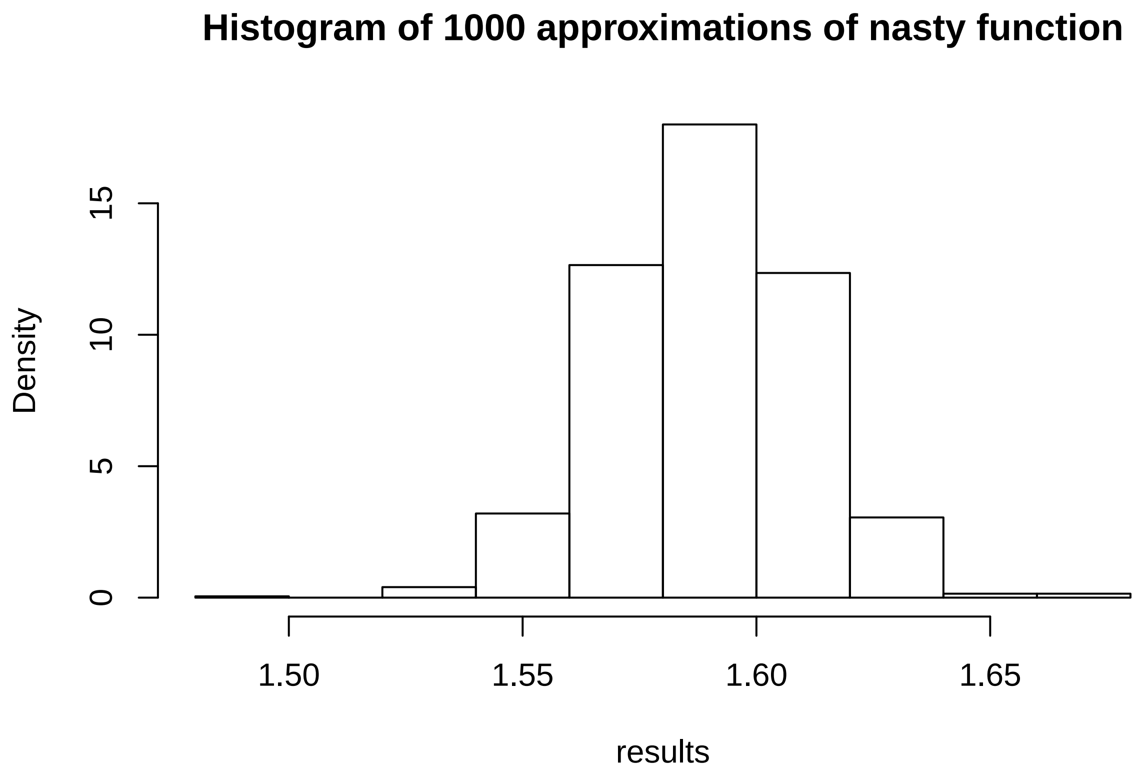
```
## [1] 1.56836
```

Note that the answer above is very similar to that which we found using the integrate function. We can increase n to get better and better approximations.

What happens if we repeat this process many times - could you predict what will happen? Let's code it up and see for ourselves.

```
results <- c()
for(i in 1:1000){
  u <- runif(n, min = 0, max = 1)
  results[i] <- mean(nasty_function(u))
}

hist(results, prob = TRUE, main = "Histogram of 1000 approximations of nasty function")
```



We see the familiar gaussian distribution centered around the value of the integral. Of course, this shouldn't surprise you - it is an instance of the Central Limit Theorem in action.

Problem 2

What if I asked you now to sample from a normal(0,1) distribution instead of uniform? Can you calculate the following integral?

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x^5 \exp(-\frac{1}{2}x^2) dx \tag{3}$$

(if you have a keen eye you will have noticed that this is the expectation of a standard normal random variable raised to the power of 5). Let's see what the value of this integral is.

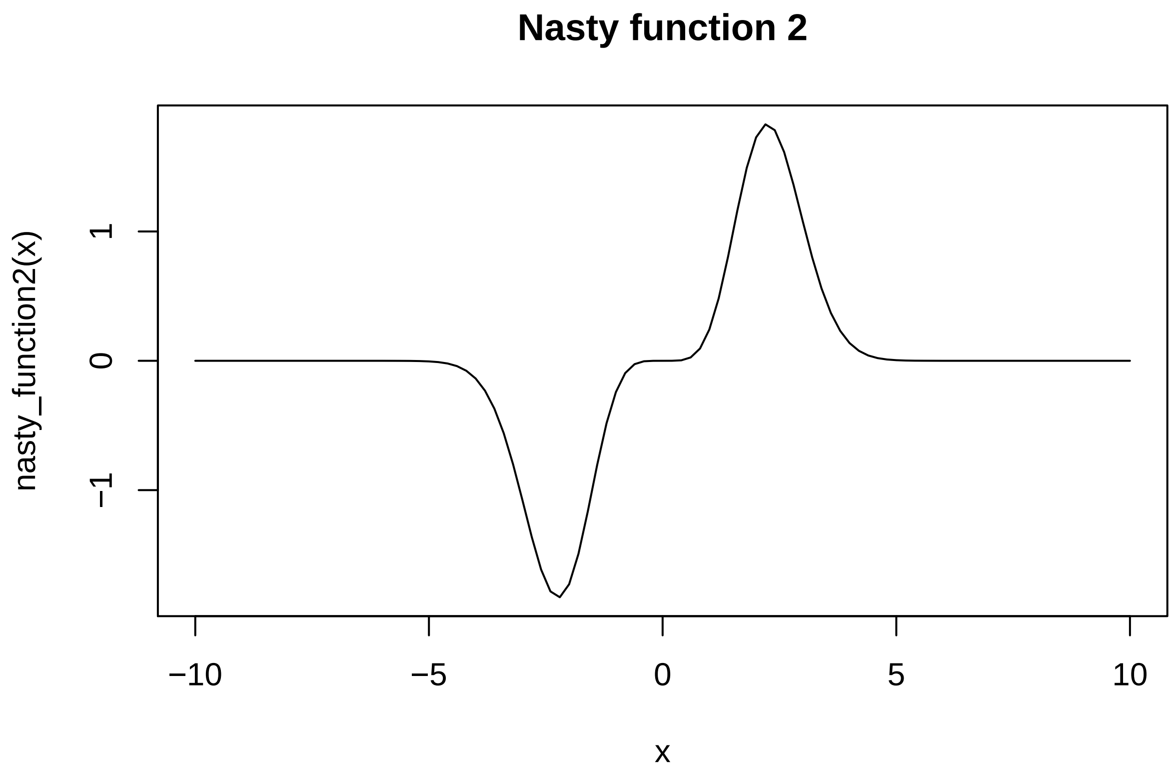
```
nasty_function2 <- function(x){
  (1/sqrt(2*pi)) * exp(-0.5 * x^2) * (x^5)
}

integrate(nasty_function2, lower = -Inf, upper = Inf)
```

```
## 0 with absolute error < 0
```

As before, let's take a look at the plot of this nasty function:

```
curve(nasty_function2, -10, 10, main = "Nasty function 2")
```



Now, let $X_1, ..., X_n \overset{iid}{\sim} N(0,1)$, where $N(0,1)$ is the standard normal distribution. Define $g(x) = x^5$, then by the law of large numbers, we get

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{p} E_Z[g(X)],$$

where E_Z is the expectation with respect to the standard normal distribution. Now, let $f(x)$ denote the pdf of a $N(0,1)$ distribution, and note that

$$E_Z[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x^5 \exp(-\frac{1}{2}x^2)dx$$

is exactly what we are after. **Remark:** Please notice that we defined $g(x)$ strategically so that the expectation would equal our desired integral. We also strategicaly chose to sample from $N(0,1)$ knowing that the density looks very similar to what we have in the integral.

To summarize, we have:

1. Step 1: Generate $X_1, ..., X_n \overset{iid}{\sim} N(0,1)$ with large n.
2. Step 2: Compute $g(X_i) = X_i^5$ for each X_i.
3. Step 3: Compute the mean of the $g(X_i)$'s. The mean will give us an approximation to the integral we seek.

The corresponding code is very simple:

```r
n <- 10000
u <- rnorm(n, mean = 0, sd = 1)
mean(u^5)
```

```
## [1] -0.1865935
```

What happens if we repeat this process over and over again? Once again, we should see a bell-shaped curve as a result of the CLT.

```r
results <- c()
for(i in 1:1000){
  u <- rnorm(n, mean = 0, sd = 1)
  results[i] <- mean(u^5)
}

hist(results, prob = TRUE, main = "Histogram of 1000 approximations of nasty function 2")
```